AD-A262 890

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

2. REPORT DATE	3. REPORT TYPE AND DATES COVERED FINAL/01 SEP 89 TO 23 FEB 92
4. TITLE AND SUBTITLE THEORY & APPLICATIONS OF NEURAL NETWORKS (U)	5. FUNDING NUMBERS 7013/00/DARPA 61101E
6. AUTHOR(S) Professor Roger Brockett	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Harvard University Applied Sciences Cambridge MA 02138	8. PERFORMING ORGANIZATION REPORT NUMBER AFOSR-TR- 5
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR/NM 110 DUNCAN AVE, SUITE B115 BOLLING AFB DC 20332-0001	10. SPONSORING/MONITORING AGENCY REPORT NUMBER AFOSR-89-0506
11. SUPPLEMENTARY NOTES	
12a. DISTRIBUTION/AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED	12b. DISTRIBUTION CODE UL

DTIC ELECTE APR 01 1993 S E D

13. ABSTRACT (Maximum 200 words)

One of the main ideas underlying the interest in neural computing is that it may be **possible to develop new computational** paradigms that will make important aspects of **programming both simple and more** robust. The means for doing so usually involves usually involves **setting** up some "universal" difference or differential equation whose trajectories **define rules** for solving problems in curve fitting, interpolation, etc. The work has **addressed the** use of analog computation methods for optimization as well as sorting, quantizing, etc.. Using a simple, but powerful, mathematical model they have shown, how basic **subsystems** can provide the building blocks that are capable of accounting for the operations that they see being performed by biological and digital computers. More specifically, they have shown that a certain class oof gradient flows on the n dimensional orthogonal group generates effective means for solving a variety of combinatorial and linear algebra problems of the type that shows up in the neural network literature. A key idea here is that of an adaptive subspace filter - a general model for nonlinear filtering of the type seen in various cognitive applications. This model not only allows one to study global convergence in a precise way, but it allows one to make analytical predictions about the speed of convergence which then can be compared with the performance of natural systems. They have shown that some of the earlier analog models for sorting can be interpreted as conditions density propagators

93-06633

14. SUBJECT TERMS			15. NUMBER OF PAGES 5
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAR(SAME AS REPORT)

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

AFOSR. 98-0500

Theory and Applications of Neural Networks

Final Technical Report. 0-01-89 to 2-29-92.

Co-Investigators: R. Brockett, L. Valiant, R. Westervelt and A. Yuille

This report is organized into four main sections. Each section gives a summary of our work performed in the area. Section (2) describes dynamical systems for analog computation with applications to optimization, adaptive filtering, and robust coding respecting the practical constriction of limited dynamic range. In Section (3) we analysed discrete neural net models to suggest practical learning algorithms. Section (4) analyzes the convergence and stability of a new class of clocked neural network models. Section (5) is based on the use of ideas from statistical physics in modelling, unification and algorithm generation.

0.1 Dynamical Systems for Analog Computation

One of the main ideas underlying the interest in neural computing is that it may be possible to develop new computational paradigms that will make important aspects of programming both simple and more robust. The means for doing so usually involves setting up some "universal" difference or differential equation whose trajectories define rules for solving problems in curve fitting, interpolation, etc. Our work [1-6] has addressed the use of analog computation methods for optimization as well as sorting, quantizing, etc. Using a simple, but powerful, mathematical model we have shown, how basic subsystems can provide the building blocks that are capable of accounting for the operations that we see being performed by biological and digital computers. More specifically, in our papers [1],[2] we have shown that a certain class of gradient flows on the *n dimensional orthogonal group generates* effective means for solving a variety of combinatorial and linear algebra problems of the type that shows up in the neural network literature. A key idea here is that of an adaptive subspace filter - a general model for nonlinear filtering of the type seen in various cognitive applications. This model not only allows one to study global convergence in a precise way, but it allows one to make analytical predictions about the speed of convergence which can then be compared with the performance of natural systems. We have shown [1] that some of our earlier analog models for sorting can be interpreted as conditional density propagators. This is important because it shows that, in a certain probabilistic sense, these are the best algorithms for doing the task in question. We have also further illuminated the connection between this work and the Toda lattice equations [4],[5],[6] and in that way have been able to provide the explicit expressions for trajectories associated with some systems of this type.

References

1. "An Estimation Theoretic Basis for the Design of Sorting and Classification Networks", R. Brockett. in *Neural Networks*, (R. Mamone and J. Zivi, Eds) Academic Press, pp. 23-41. 1991.

2. "Dynamical Systems That Learn Subspaces," R. Brockett. In *Mathematical System Theory: The Influence of R. E. Kalman*, (A.C. Antoulas, Ed.) Springer Verlag, Berlin, pp. 579-592. 1991.

3. "Completely Integrable Gradient Flows", A.M. Bloch, R. Brockett and T. Ratiu. *Communications in Mathematical Physics*, Vol. 147, pp. 57–74. 1992.

4. "Toda Flows, Inverse Spectral Transform and Realization Theory," R. Brockett and L.E. Faybusovich. *Systems and Control Letters*, Vol. 16, No. 2, pp. 79-88. 1991.

5. "Sorting With the Dispersionless Limit of the Toda Lattice," R. Brockett and A.M. Bloch, in *Hamiltonian Systems, Transformation Groups and Spectral Transform Methods*, CRM, (J. Harnad and J.E. Marsden, Eds.) Université de Montréal, Montréal, Canada pp 103-11 1992.

6. "Spectral Equations for the Long Wave Limit of the Toda Lattice Equations", A.M Bloch, R. Brockett, Y. Kodama and T. Ratiu, in *Hamiltonian Systems, Transformation Groups and Spectral Transform Methods*, CRM, (J. Harnad and J.E. Marsdon Eds.) Université de Montréal, Montréal, Canada. pp. 97-102. 1992.

0.2 Learning Theory and Discrete Neural Nets

We have been fortunate to have had two excellent postdoctoral fellows, Nick Littlestone and Robert Schapire. They are among the leading contributors to computational learning theory. Their work on this grant is described in the six publications below. Each paper addresses in some way the problem of how a function can be learned from seeing examples and counterexamples of it using only feasible computational resources.

A brief summary of some of these papers is as follows: (i) Successful studies of the learning curves of neural nets exist in two very different frameworks. One is the Vapnik-Cheronenkis dimension from statistics, and the other is the language of statistical physics. In reference [1] a very elegant unification of these two approaches is presented. (ii) A few years ago Littlestone discovered a learning algorithm for perceptrons that is as simple as the classical one, but can be shown to outperform it greatly in the case that a large number of the dimensions are irrelevant. In most realistic settings of learning this is the case since the relevant attributes cannot be identified *a priori* in general. In reference [2] Littlestone demonstrates further properties of his algorithm, in the important case that some errors in the data have to be allowed for. (iii) On the subject of coping with large numbers of attributes that are irrelevant but not identified as such, in reference [3] the first general transformations are presented for translating any one from a wide class of learning algorithms to one that is attribute efficient. (iv) In [4] Schapire gives one of the most general positive analytic results known for learning a class of functions. Reference [5] introduces novel techniques to prove that certain classes that are not learnable in the general case are learnable for restricted input distributions.

References

1. "Bounds are the sample complexity of Baeysian learning using information theory and the VC Dimension". D. Haussler, M. Kearns and R. Schapire. *Proc. 4th Workshop on Computational Learning Theory* (1991) 61-74.

2. "Redundant noisy attributes, attribute errors, and linear-threshold learning using winnow". N. Littlestone. *Proc. 4th Workshop on Computational Learning Theory* (1991) 147-156.

3. "Learning in the presence of finitely or infinitely many irrelevant attributes". A. Blum, L. Hellerstein and N. Littlestone. *Proc. 4th Workshop on Computational Learning Theory*, (1991) 157-166.

4. "Learning probabilistic read-once formulas on product distributions". R.E. Schapire. *Proc. 4th Workshop on Computational Learning Theory* (1991) 184-198.

5. "Exact identification of read-once formulas using fixed points of amplification functions". S. A. Goldman, M.J. Kearns and R.E. Schapire. *Proc. 31st IEEE Symp. on Foundations of Computer Science* (1990) 193-202.

0.3 Planar Neural Networks for Vision and Sound Processing.

We conducted a systematic study of the stability of neural networks with feedback, including the delays which occur in real implementations [1]. The goal of an architecture which could be stably implemented in hardware led us to develop the clocked neural network [2,3] in which all neurons are updated synchronously on a clock pulse. The effects of clocking are analogous to those for digital computers: synchronous update helps to stabilize the network by eliminating timing ambiguities caused by varying signal paths or neuron delays. One of our most significant achievements is the proof of a global stability criterion [2] for clocked neural networks with arbitrary symmetric interconnections. The conditions of the proof are sufficiently general that the results guarantee stability for real implementations of clocked networks in hardware: for example, the neuron transfer characteristics are continuous and can differ from neuron to neuron. We have applied this approach to redesign feedback associative memories of the type originally considered by Hopfield as clocked neural networks, and have computed "phase diagrams" which specify the stable operating region in terms of the neuron gain and the number of stored memories [4,5]. We have also computed the number of small spurious attractors which can prevent the state of the system from reaching a stored memory [6,7,8]. The sum of this work constitutes a solution to the mathematical problem of stability in many types of feedback neural networks with a wide variety of interconnection topologies and interconnection strengths.

References

1. "Stability of Analog Neural Networks with Delay", Marcus, C.M., and R.M Westervelt Phys. Rev. A **39**, 347. 1989

2. "Dynamics of Iterated Map Neural Networks", Marcus, C.M., and R.M.Westervelt. Phys Rev. A **40**, 501. 1989

3. "Dynamics of Analog Neural Networks with Time Delay", Marcus, C.M., and R.M Westervelt in *Advances in Neural Information Processing Systems*, Edited by D. S. Touretzky (Morgan Kaufman, San Mateo, California), p. 568. 1989

4. "Associative Memory in an Analog Iterated Map Neural Network", Marcus, C.M., F.R Waugh and R.M.Westervelt. Phys. Rev. A **41**, 3355. 1990

5. "Stability and Convergence of Analog Neural Networks with Multiple Time Step Parallel Dynamics", Marcus, D.M., and R.M.Westervelt. Phys. Rev. A, **42**, 2410. 1990

6. "Nonlinear Dynamics of Analog Associative Memory Neural Networks", Waugh, F.R., C.M.Marcus. and R.M.Westervelt,in *Proc. Int. Joint Conf. on Neural Networks, Washington, DC, 1990*. Ed. Maureen Caudill (Lawrence Erlbaum Assoc., Hillsdale, NJ) vol. 1, p. 321. 1990.

7. "Fixed-point Attractors in Analog Neural Computation", Waugh, F.R., C.M.Marcus and R.M.Westervelt.Phys. Rev. Lett. **64**, 1990.

8. "Counting Fixed-point Attractors in an Analog Associative Memory", Waugh, F.R., C.M. Marcus and R. M. Westervelt. Phys. Rev. A **43**, 3131. 1991.

9. "Connection Topology and Dynamics in Lateral Inhibition Networks" Marcus, C.M., F.R.Waugh. and R.M.Westervelt. In *Advances in Neural Information Processing Systems 3* (Morgan Kaufman, San Mateo) p. 98. 1991.

0.4 Statistical Physics for Modelling, Optimization and Unification

Our research supported by this grant concentrated on five areas based on the statistical physics approach to neural networks: (i) optimization for combinatorial problems using deterministic annealing, (ii) using deformable templates for high energy particle detection, (iii) modelling binocular stereo visual perception and relating it to psychophysics, (iv) modelling the development of the visual cortex, and (v) enabling these statistical physics models to incorporate techniques used in statistics.

There are three main themes in this work: (a) using energy functions and the Gibbs distribution to model problems in terms of finding the optimal statistical estimators, (b) using deterministic annealing to obtain the estimators, and (c) unification by using techniques from statistical physics to show relationships between different theories. These themes are emphasized in a short review paper [1].

In the work on optimization we focussed initially on the assignment problem [2,3] though we are currently generalizing this work. We proved that two dynamical systems using deterministic annealing were guaranteed to converge to the optimal solution and gave bounds on the convergence times of these systems. We also found intriguing relations between these dynamical systems and more traditional approaches to these problems such as the auction algorithm, interior point methods and linear programming with barrier functions [11].

We applied deformable templates in a statistical physics framework for the detection of particles in high energy physics experiments [4,5]. The resulting system, using deterministic annealing, performed well on simulated 2D data given to us by the LEP lab in CERN. This was extended to work on 3D data [9].

The work on stereo [6] proposed a general formulation of stereo, showed that we could use statistical physics techniques to relate previous theories to this framework, proposed deterministic annealing as an optimization strategy, and finally showed that the resulting theory was consistent with a number of psychophysical experiments.

By embedding models of self-organization of the visual cortex within a statistical physics framework we were able to show [7] precise, and hitherto unexpected, relations between existing theories of ocularity and the spatial organization of orientation. This emphasized the importance of the optimization criteria over the mechanism proposed for minimizing it. Small variations in the criteria can lead to very different experimental predictions.

Other work [8] investigated a toy theory of stereo transparency perception, showed that the statistical physics approach could incorporate techniques from robust statistics, and that the Hough/Radon transform could appear as a special case in the zero temperature limit. We also [10] suceeded in determining the phase transitions for a class of texture synthesis models.

References

1. "Statistical Physics for Modelling and Optimization". A.L. Yuille and D. Geiger. Harvard Robotics Laboratory Technical Report 91-2. 1991

2. "The Invisible Hand Algorithm: Solving the Assignment Problem with Statistical Physics" J. Kosowsky and A.L. Yuille. *Proceedings IJCNN*. Seattle. WA. July. 1991. See also Harvard Robotics Laboratory Technical Report 91-1. 1991.

3. "The Invisible Hand Algorithm: Time Convergence and Temperature Tracking." A.L. Yuille, and J.J. Kosowsky. Harvard Robotics Laboratory Technical Report 91-10. 1991.

4. "Deformable Templates for Particle Tracking" A.L. Yuille, K. Honda and C. Petersen. *Proceedings IJCNN*. Seattle, WA. July, 1991. See also: Harvard Robotics Laboratory Technical Report. 90-8. 1990.

5. "Deformable Templates, robust statistics, and Hough transforms". A.L. Yuille, C. Peterson and K. Honda. *Proceedings SPIE Geometric Methods in Computer Vision*, San Diego, CA. July, 1991.

6. "Stereo Integration, Mean Field and Psychophysics" A.L. Yuille, D. Geiger and H. H. Bulthoff *NETWORK*. 2. pp 432-442. 1991.

7. "Dimension Reduction, Generalized Deformable Models and the Development of Occularity". A.L. Yuille, J. Kolodny and C. Lee. *Proceedings IJCNN*. Seattle, WA. July, 1991. See also Harvard Robotics Laboratory Technical Report 91-3. 1991.

8. "Robust Statistics, Transparency and Correspondence", A.L. Yuille, T. Yang and D. Geiger. Harvard Robotics Laboratory Technical Report 90-7. 1990.

9. "Track finding with deformable templates – the elastic arms approach". M. Ohlsson, C. Peterson and A.L. Yuille. *Computer Physics Communications*. 71, pp 77-98. 1992.

10. "Mean-Field Theory and Phase Transitions for Grayscale Texture Synthesis". A. L. Yuille and I.M. Elfadel. *Proceedings of 1992 Conference on Information Sciences and Systems*., pp 350-355. Princeton. 1992.

11. "Statistical Physics Algorithms that Converge". A.L. Yuille and J.J. Kosowsky. Harvard Robotics Laboratory Technical Report 92-7. 1992.